

Estimation de cartes de disparité basée sur l’algorithme revisité d’appariement de blocs pour la compression d’images stéréoscopiques

Aysha KADAIKAR, Gabriel DAUPHIN, Anissa MOKRAOUI

Laboratoire de Traitement et Transport de l’Information
Institut Galilée, Université Paris 13, Sorbonne Paris Cité 99 avenue Jean-Baptiste Clément, 93430 Villetaneuse, France
kadaikar@univ-paris13.fr, gabriel.dauphin@univ-paris13.fr
anissa.mokraoui@univ-paris13.fr

Résumé – Cet article traite du problème de l’estimation de cartes de disparité en vue de la compression d’images stéréoscopiques. L’algorithme classique d’appariement de blocs associe à chaque bloc de la vue à prédire une disparité. Ces disparités sont choisies de façon à minimiser la distorsion globale évaluée entre la vue originale et sa prédiction. Pour une distorsion minimale, le coût binaire de la carte estimée peut cependant s’avérer élevé. Dans le but d’assurer un bon compromis débit-distorsion, nous proposons une méthode d’estimation de carte basée sur une métrique conjointe entropie-distorsion. Partant d’une carte de disparité de référence, celle estimée par l’algorithme classique d’appariement de blocs, nous proposons de modifier au fur et à mesure cette carte tant que les modifications envisagées réduisent davantage la métrique conjointe. La nouvelle carte estimée est ensuite traitée autant de fois qu’il est possible d’obtenir une diminution de la métrique conjointe. De plus, une stratégie de mise à jour de la métrique conjointe est proposée de façon à ne pas affecter la complexité de calcul de notre algorithme. Les simulations, réalisées sur la base d’images stéréoscopiques Middlebury, montrent une amélioration significative en termes de débit-distorsion comparée à l’algorithme classique d’appariement de blocs.

Abstract – This paper deals with the problem of estimating disparity maps for stereoscopic image compression. The traditional block-matching algorithm associates a disparity to each block of the view to be predicted. These disparities are selected so that they minimize the distortion computed between the original and the predicted view. However the binary cost involved by the resulting disparity map minimizing the distortion may be high. In order to ensure a good bitrate-distortion trade-off, we propose a map estimation method based on a joint entropy-distortion metric. The proposed approach relies on a reference solution, initially taken as the disparity map computed by the traditional block-matching algorithm, which is modified successively as long as the proposed changes reduce the entropy-distortion metric. The new estimated disparity map is then processed again as long as a decrease in the joint metric is achieved. In addition, an update strategy of the joint metric is proposed so as the computational complexity of the algorithm is not affected. Simulations, conducted on stereoscopic images from Middlebury dataset, show that significant gains are achieved in terms of rate-distortion compared to the traditional block-matching algorithm.

1 Introduction

Ces dernières décennies ont vu émerger un bon nombre d’applications offrant à l’utilisateur une impression de profondeur, citons par exemple la TV-3D, les jeux vidéos immersifs ou encore la visioconférence en 3D [1, 2]. Cette impression de profondeur n’est rien d’autre qu’une illusion d’optique lorsque deux vues d’une même scène sont présentées à un observateur. Plus un objet d’une scène apparaît distant (impression de grande profondeur), plus il est repéré dans les vues à des endroits proches. L’ensemble des déplacements spatiaux des objets présents dans les deux vues forme une carte de disparité. Etant donné que les deux vues sont ressemblantes, trouver une technique de codage pour exploiter les redondances inter-vues est aujourd’hui un sujet de recherche d’autant plus important que les applications récentes semblent présenter un intérêt croissant.

La compression d’images stéréoscopiques s’appuie assez souvent sur un schéma de codage par compensation de disparité. L’une des deux vues, par exemple la gauche, est choisie comme

vue de base pour être comprimée indépendamment de la vue droite. Une carte de disparité est estimée puis comprimée selon un codage entropique [3, 4]. La vue droite est alors prédite en utilisant la carte estimée ainsi que la vue de base. Une image résiduelle est calculée entre l’image droite originale et sa prédiction avant d’être comprimée. Il s’en suit alors que la carte estimée joue un rôle primordial sur les performances de la méthode de compression.

L’Algorithme de Mise en correspondance de Blocs (AMB) (ou d’appariement de blocs) est assez souvent exploité pour estimer la carte de disparité. Cette dernière, constante par bloc (c.a.d. une même disparité pour tous les pixels d’un bloc de l’image), permet de réduire de manière importante le coût binaire comparé aux cartes denses (c.a.d. une disparité par pixel). L’AMB sélectionne les disparités qui minimisent une métrique de dissimilarité calculée lors de l’appariement des blocs, telles que la somme des carrés des différences ou la somme des différences absolues évaluées entre les intensités des blocs à apparier [5]. Pour réduire le coût du codage de la carte de

disparité, l'AMB ne peut qu'ajuster la taille des blocs qui cependant peut détériorer la qualité de la vue reconstruite.

Pour améliorer les performances de l'AMB, nous proposons d'exploiter une métrique conjointe entropie-distorsion sur laquelle s'appuie l'estimation de la carte de disparité. Pour satisfaire ce compromis, une solution sous-optimale est proposée par l'algorithme appelé R-algorithme (R pour Référence) que nous avons développé. Ce dernier initie son traitement sur la carte de disparité estimée par l'AMB, dite carte de référence. Cette carte est analysée au fur et à mesure que l'algorithme réexamine les blocs à apparier. Elle est mise à jour uniquement lorsque la modification envisagée entraîne une réduction de la métrique conjointe. L'ensemble du traitement est itéré sur la nouvelle carte jusqu'à ce qu'il ne soit plus possible de réduire la métrique conjointe.

Cet article est organisée comme suit. La section suivante énonce le problème de minimisation de la métrique conjointe entropie-distorsion sur laquelle s'appuie l'estimation des cartes de disparité des images stéréoscopiques. La section 3 décrit ensuite la solution sous-optimale proposée par le R-algorithme. La section 4 discute des résultats de simulation présentés. La dernière section conclut notre travail.

2 Estimation de la carte de disparité

Cette section introduit tout d'abord quelques notations puis présente le problème de minimisation que notre algorithme cherchera à résoudre en section 3 en vue d'estimer la carte de disparité associée à l'image stéréoscopique.

2.1 Hypothèses et notations

Nous supposons que les deux vues de l'image stéréoscopique sont rectifiées. L'appariement des blocs s'effectue alors en supposant uniquement un déplacement horizontal. La vue gauche est choisie comme étant la vue de base (habituellement appelée référence) tandis que la vue droite correspond à la vue prédite.

Les vues gauche (I_g) et droite (I_d) de l'image stéréoscopique sont de taille $K \times L$ pixels. Chaque vue est divisée en $X \times Y$ blocs distincts, chacun de taille $N_X \times N_Y$ pixels. Chaque bloc est repéré par ses coordonnées (i_b, j_b) parmi l'ensemble des blocs $X \times Y$, et peut être par ailleurs représenté par les coordonnées du pixel situé au coin gauche supérieur du bloc $(i_p, j_p) : i_p = i_b \times N_X$ et $j_p = j_b \times N_Y$.

$I_g(i_p, j_p)$ et $I_d(i_p, j_p)$ représentent l'intensité du pixel de coordonnées (i_p, j_p) respectivement dans la vue gauche et la vue droite. On note $d(i_b, j_b)$ la disparité associée au bloc de coordonnée (i_b, j_b) . Cette disparité, entière ou non-entière, est définie sur l'intervalle $W = \{w_{min}, w_{min} + 1/\alpha, \dots, w_{max} - 1/\alpha, w_{max}\}$, où $\frac{1}{\alpha}$ (avec $\alpha \in \{1, 2, 4, 8\}$) apporte une précision à la disparité ; w_{min} et w_{max} délimitent l'espace de recherche. Le nombre de disparités possibles est alors égal à $N_\alpha = (w_{max} - w_{min}) \times \alpha + 1$. On note $\hat{I}_d(i_p + u, j_p + v)$ l'intensité prédite du pixel de la vue droite sachant que la disparité associée au bloc (i_b, j_b) est $d(i_b, j_b)$. Dans la cas où la disparité est non-

entière, une interpolation sous-pixellique est requise. La carte de disparité recherchée est notée $\mathbf{d} = [d(i_b, j_b)]$ avec $i_b = 0, \dots, X - 1; j_b = 0, \dots, Y - 1$.

2.2 Métrique conjointe entropie-distorsion

Cette section propose d'estimer la carte de disparité qui offre le meilleur compromis débit-distorsion.

La distorsion globale, $E_G(\mathbf{d})$, induite par la vue reconstruite est donnée par la somme des distorsions locales :

$$E_G(\mathbf{d}) = \sum_{i_b=0}^{X-1} \sum_{j_b=0}^{Y-1} E_L((i_b, j_b), d(i_b, j_b)), \quad (1)$$

où les distorsions locales sont calculées comme suit :

$$E_L((i_b, j_b), d(i_b, j_b)) = \sum_{u=0}^{N_X-1} \sum_{v=0}^{N_Y-1} (\hat{I}_d(i_p + u, j_p + v) - I_d(i_p + u, j_p + v))^2. \quad (2)$$

Le coût binaire associé à la carte de disparité est approximé par l'entropie donnée par la relation suivante :

$$H(\mathbf{d}) = - \sum_{w \in W} \frac{V(\mathbf{d}, w)}{T} \log_2 \left(\frac{V(\mathbf{d}, w)}{T} \right), \quad (3)$$

où $V(\mathbf{d}, w)$ représente le nombre d'occurrence de la disparité w dans la carte de disparité \mathbf{d} . T désigne le nombre total de blocs dans l'image ($T = X \times Y$). Le ratio $\frac{V(\mathbf{d}, w)}{T}$ correspond à la probabilité empirique d'apparition de la disparité w dans la carte de disparité \mathbf{d} .

Le problème d'optimisation du compromis entropie-distorsion est alors formalisée par une minimisation Lagrangienne :

$$J(\lambda, \mathbf{d}) = E_G(\mathbf{d}) + \lambda H(\mathbf{d}), \quad (4)$$

où λ est le multiplicateur de Lagrange.

Le coût $J(\lambda, \mathbf{d})$ correspond à une métrique conjointe entropie-distorsion.

3 Description du R-algorithme proposé

Cette section propose une solution sous-optimale au problème de minimisation de la métrique entropie-distorsion ($J(\lambda, \mathbf{d})$) introduit à la section précédente. L'algorithme d'optimisation, appelé R-algorithme (R pour Référence), initie son traitement sur la carte de disparité fournie par l'AMB. Cette carte est considérée comme une carte de référence initiale. Elle est modifiée progressivement autant de fois qu'une réduction de la métrique globale est observée. Le processus complet est expliqué ci-dessous.

Avant de décrire le R-algorithme, introduisons quelques notations utiles pour la suite. L'algorithme traite les blocs de l'image de base à apparier de manière séquentielle selon un balayage en ligne. Pour des raisons de simplicité, les coordonnées d'un bloc (i_b, j_b) sont remplacées par une coordonnée unidimensionnelle, notée t , comme suit :

$$t = i_b \times Y + j_b + 1. \quad (5)$$

$d(i_b, j_b)$ est alors remplacée par d_t et $E_L((i_b, j_b), d(i_b, j_b))$ par $E_L(t, d_t)$. La carte de disparité initiale, dite de référence, est notée :

$$\mathbf{d}^R = [d_1^R, d_2^R, \dots, d_t^R, \dots, d_T^R], \quad (6)$$

où d_t^R représente la disparité associée au t -ème bloc.

3.1 Principe du R-algorithme

Supposons que notre R-algorithme ait traité les $(t - 1)$ premiers blocs. La carte de disparité mise à jour correspond alors à :

$$\mathbf{d}^{t-1} = [d_1, d_2, \dots, d_{t-1}, d_t^R, \dots, d_T^R], \quad (7)$$

où d_1, d_2, \dots, d_{t-1} représentent les disparités sélectionnées par le R-algorithme et d_t^R, \dots, d_T^R sont les disparités de la carte de référence.

Le traitement du t -ème bloc consiste à permuter la disparité d_t^R avec toutes les autres disparités possibles w définies dans $W \setminus \{d_t^R\}$. $N_\alpha - 1$ nouvelles cartes de disparité sont alors construites :

$$\mathbf{d}^t(w) = \{d_1, d_2, \dots, d_{t-1}, w, d_{t+1}^R, \dots, d_T^R\}. \quad (8)$$

La disparité à retenir correspond à celle qui vérifie les conditions suivantes :

$$d_t := w \text{ si } \begin{cases} w = \operatorname{argmin}_{w' \in W \setminus \{d_t^R\}} J(\lambda, \mathbf{d}^t(w')) \\ J(\lambda, \mathbf{d}^t(w)) < J(\lambda, \mathbf{d}_t^R) \end{cases}, \quad (9)$$

à savoir celle qui diminue la métrique J donnée par (4).

Le processus est itéré jusqu'à la lecture du dernier bloc (c.a.d le T -ème bloc). Une nouvelle carte de disparité \mathbf{d}^T est alors construite. Cette dernière est considérée à son tour comme une nouvelle carte de référence \mathbf{d}^R . Le R-algorithme réitère alors l'ensemble du processus décrit ci-dessus jusqu'à ne plus pouvoir obtenir de diminution de la métrique conjointe.

3.2 Mise à jour de la métrique entropie-distorsion

Pour réduire la complexité de calcul liée notamment à la mise à jour de la métrique conjointe entropie-distorsion, nous proposons des équations récursives de façon à éviter de reprendre tous les calculs.

Supposons que d_t^R , une des disparités de la carte \mathbf{d}^{t-1} , est remplacée par w pour construire la nouvelle carte $\mathbf{d}^t(w)$. La mise à jour de la distorsion globale E_G , définie dans (1) avec la notation de (5), est alors donnée par l'équation recursive suivante :

$$E_G(\mathbf{d}^t(w)) = E_G(\mathbf{d}^{t-1}) - E_L(t, d_t^R) + E_L(t, w), \quad (10)$$

où $E_L(t, d_t^R)$ et $E_L(t, w)$ sont définies dans (2) et (5).

L'entropie H , définie dans (3) avec la notation de (5), est alors mise à jour par la relation recursive suivante :

$$\begin{aligned} H(\mathbf{d}^t(w)) &= H(\mathbf{d}^{t-1}) \\ &+ \left(\frac{V(\mathbf{d}^{t-1}, d_t^R)}{T} \right) \log_2 \left(\frac{V(\mathbf{d}^{t-1}, d_t^R)}{T} \right) \\ &- \left(\frac{V(\mathbf{d}^{t-1}, d_t^R) - 1}{T} \right) \log_2 \left(\frac{V(\mathbf{d}^{t-1}, d_t^R) - 1}{T} \right) \\ &+ \left(\frac{V(\mathbf{d}^{t-1}, w)}{T} \right) \log_2 \left(\frac{V(\mathbf{d}^{t-1}, w)}{T} \right) \\ &- \left(\frac{V(\mathbf{d}^{t-1}, w) + 1}{T} \right) \log_2 \left(\frac{V(\mathbf{d}^{t-1}, w) + 1}{T} \right). \end{aligned} \quad (11)$$

Notons que $V(\mathbf{d}^t(w), w')$ est en effet identique à $V(\mathbf{d}^{t-1}, w')$ à ceci près qu'il est décrétement d'une unité lorsque $w' = d_t^R$ et incrémenté d'une unité lorsque $w' = w$. La mise à jour de J définie dans (4) résulte alors de (10) et (11).

4 Résultats expérimentaux

Les simulations ont été réalisées sur les images stéréoscopiques issues des bases Middlebury et Deimos [6, 7]. La vue de base correspond à la vue gauche. La vue droite est prédite à partir de la carte de disparité estimée. Les performances du R-algorithme sont comparées à celles de l'AMB. La qualité des images prédites est mesurée en termes de PSNR (Peak Signal to Noise Ratio, mesurée en dB). Le coût binaire associé à la carte de disparité est approximé par l'entropie (mesurée en bits par pixel (bpp)).

La Figure 1 compare les performances, en termes de débit-distorsion (dB vs bpp), de notre algorithme à celles de l'AMB sur l'image stéréoscopique "Barn2" de taille 381×430 pixels pour différentes tailles de bloc : de 4×4 jusqu'à 16×16 pixels. Les disparités sont définies sur $[-30, -30+1/2, \dots, 29 + 1/2]$. La vue de base a été interpolée au demi-pixel en exploitant les filtres utilisés dans la norme H.264 [8].

La courbe en pointillée qui relie les symboles "circulaires" concerne l'AMB pour différentes tailles de bloc. Les courbes en trait plein (en rouge), tracées pour un ensemble de valeurs de λ positives, illustrent les performances de notre R-algorithme. Chaque courbe correspond à une taille de bloc choisie. Les simulations confirment ici qu'au sein de chaque courbe, lorsque λ augmente, le débit et le PSNR ont tendance à diminuer. L'extrémité droite de chacune de ces courbes coïncident avec un des cercles obtenu par l'AMB. En effet lorsque $\lambda = 0$, le R-algorithme est identique à l'AMB.

Ces courbes montrent que notre R-algorithme réduit significativement le débit binaire au prix d'une petite baisse de la qualité de prédiction. En effet pour des blocs de taille 4×4 , l'AMB permet de prédire la vue droite avec une qualité de 34.51dB et un débit binaire de 0.247bpp ; avec un choix particulier de λ et des blocs de même taille que l'AMB, le R-algorithme réduit la qualité de prédiction de 0.3dB pour un débit binaire qui vaut seulement 0.088bpp.

La Figure 1 montre aussi que vis-à-vis de l'AMB, le R-algorithme permet d'avoir une amélioration significative de la qualité de prédiction pour un même débit en utilisant des tailles de blocs inférieures et un choix judicieux de la valeur de λ . Ainsi avec des blocs de taille 7×7 , l'AMB offre une qualité de reconstruction de 33.47dB avec un débit de 0.075bpp. Le R-algorithme, avec des blocs de taille 5×5 et une valeur particulière de λ , offre pour un débit similaire (0.073bpp) une qualité de reconstruction de 34.09dB soit un gain de 0.6dB en notre faveur.

La Figure 2 correspond à la vue droite originale de l'image stéréoscopique "Barn2". La Figure 3 montre deux agrandissements de la zone marquée par le rectangle sur la vue origi-

nale donnée par la Figure 2. De gauche à droite, cette figure donne les reconstructions obtenues avec les cartes de disparité estimées par l'AMB et le R-algorithme. Ces résultats sont fournis pour des coûts binaires identiques, mais des blocs de différentes tailles : 7×7 pour l'AMB et 5×5 pour le R-algorithme. Sur l'image associée à l'AMB, nous constatons que le sommet du triangle a disparu alors qu'il apparaît parfaitement lorsque la reconstruction s'appuie sur la carte de disparité estimée par notre R-algorithme.

La Table 1 fournit des résultats obtenus sur d'autres images stéréoscopiques. Les tests ont été effectués pour des débits binaires faibles, moyens et élevés obtenus respectivement avec des blocs de taille 8×8 , 6×6 et 4×4 avec les deux algorithmes. Les disparités possibles sont définies sur $[-30, -30+1/4, \dots, 29 + 1/4]$. La vue de base a été interpolée au quart de pixel en exploitant les filtres utilisés dans la norme H.264 [8]. Comparé à l'AMB, l'analyse des résultats permet de constater que notre R-algorithme réduit significativement le débit nécessaire à l'encodage de la carte de disparité au détriment d'une baisse très légère de la qualité de la vue prédite.

5 Conclusion

Nous avons développé un algorithme d'estimation de cartes de disparité, constante par bloc, pour la compression d'images stéréoscopiques. Notre algorithme repose sur la minimisation d'une métrique conjointe entropie-distorsion. Appelé R-algorithme (R pour référence), il initie son traitement sur la carte de disparité, dite de référence, construite par l'AMB. Cette carte est modifiée tant que des réductions de la métrique conjointe sont observées. Par ailleurs, il propose une réduction de la charge de calcul lors de l'actualisation de la métrique conjointe. Bien que la solution proposée soit sous-optimale, notre R-algorithme apporte de meilleurs résultats que l'AMB. En effet, une baisse significative du débit binaire au prix d'une légère diminution de la qualité de reconstruction est observée. De plus, pour des blocs de taille inférieure à celle de l'AMB et un même débit binaire, une amélioration significative de la qualité subjective et objective de la vue reconstruite est constatée. Nous envisageons, dans des travaux futurs, d'intégrer notre algorithme dans un schéma complet de codage par compensation de disparité.

Images stéréo.	Taille du bloc	AMB		R-algorithme	
		PSNR	Débit	PSNR	Débit
Tsukuba	4x4	35.12	0.338	34.98	0.188
	6x6	32.95	0.138	32.88	0.084
	8x8	32.08	0.074	32.00	0.046
Stereo_13	4x4	30.57	0.430	30.48	0.317
	6x6	29.35	0.180	29.27	0.124
	8x8	28.88	0.097	28.80	0.065
Sawtooth	4x4	33.37	0.362	33.30	0.285
	6x6	31.79	0.154	31.74	0.126
	8x8	30.69	0.085	30.62	0.069

TABLE 1 – Comparaisons des performances débit-distorsion.

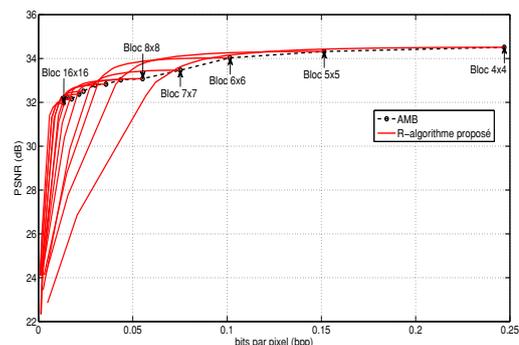


FIGURE 1 – Optimisation débit-distorsion sur "Barn2".



FIGURE 2 – Vue droite originale de "Barn2".

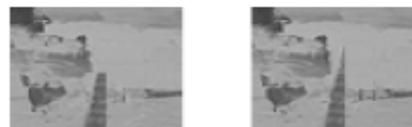


FIGURE 3 – Zoom sur la reconstruction de la zone indiquée sur la Figure 2, de gauche à droite : l'AMB ; le R-algorithme.

Références

- [1] N. A. Dodgson "Autostereoscopic 3D Displays," *IEEE Computer*, vol. 38(8), pp. 31–36, Aug. 2005.
- [2] P. Kauff and O. Schreer "An immersive 3D video-conferencing system using shared virtual team user environments," *Proceedings of the 4th international conf. on Collaborative virtual environments*, pp. 105–112, Sept. 2002.
- [3] M. G. Perkins, "Data compression of stereo pairs," *IEEE Trans. on Comm.*, vol. 40(4), pp. 684–696, Apr. 1992.
- [4] W. Woo and A. Ortega, "Stereo image compression based on disparity field segmentation," *SPIE Conf. on Visual Comm. and Image Processing*, vol. 3024, pp. 391–402, Feb. 1997.
- [5] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo," *IEEE Trans. on Pattern Anal. and Machine Intelligence, PAMI*, vol. 25(8), pp. 993–1008, Aug. 2003.
- [6] M. Klima and al. "DEIMOS-an open source image database," *Radioengineering*, vol. 20(4), pp. 1016–1023, 2011. "http://www.deimos-project.eu/tag/stereo".
- [7] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondance Algorithms," *Int. Journal of Computer Vision, IJCV*, vol. 47(1), pp. 7–42, Apr. 2002. "http://vision.middlebury.edu/stereo/data/".
- [8] I. E. Richardson "The H. 264 advanced video compression standard," *John Wiley & Sons*. 2011.